



Learning from video modeling examples: does gender matter?

Vincent Hoogerheide¹ · Sofie M. M. Loyens^{1,2} ·
Tamara van Gog^{1,3}

Received: 6 January 2015 / Accepted: 3 October 2015 / Published online: 12 October 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract Online learning from video modeling examples, in which a human model demonstrates and explains how to perform a learning task, is an effective instructional method that is increasingly used nowadays. However, model characteristics such as gender tend to differ across videos, and the model-observer similarity hypothesis suggests that such characteristics may affect learning. Therefore, this study investigated whether the effectiveness of learning how to solve a probability calculation problem from video modeling examples would vary as a function of the model's and observer's gender. In a 2 (Model: Female/Male) × 2 (Observer: Female/Male) between-subject design, 167 secondary education students learned how to solve probability calculation problems by observing video modeling examples. Results showed no effects of Model or Observer gender on learning and near transfer. Male students reported higher self-efficacy than female students. Compared to a female model, observing a male model enhanced perceived competence more from pretest to posttest, irrespective of observers' gender. Furthermore, learning from a male model was less effortful and more enjoyable for male students than for female students. These results suggest that gender of both model and observer can matter in terms of affective variables experienced during learning, and that instructional designers may want to consider this when creating (online) learning environments with video modeling examples.

Keywords Example-based learning · Multimedia learning · Modeling · Model-observer similarity · Gender

✉ Vincent Hoogerheide
hoogerheide@fsw.eur.nl

¹ Institute of Psychology, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

² Roosevelt Center for Excellence in Education, University College Roosevelt, Middelburg, The Netherlands

³ Department of Pedagogical and Educational Sciences –Education, Utrecht University, Utrecht, The Netherlands

Introduction

Students of all ages and educational levels increasingly watch instructional videos for informal learning purposes online on websites such as YouTube and Google Videos, but such videos are also increasingly used in formal learning (Lenhart 2012; Spires et al. 2012). In formal learning, online instructional videos can be consulted while making homework, or can replace activities that normally took place face to face. For instance, some educators have even argued in favor of a “flipped classroom”, which entails having learners study videos at home to free up time in school for practice and teacher support (Bergmann and Sams 2012). Various types of videos are used for both informal and formal learning purposes, such as web lectures (e.g., Day and Foley 2006; Traphagan et al. 2010), short knowledge clips (e.g., Day 2008), and how-to demonstration videos (e.g., Ayres et al. 2009). Regarding the latter, research inspired by social-cognitive theories such as social learning theory (Bandura 1977, 1986) and cognitive apprenticeship (Collins et al. 1989) has demonstrated the effectiveness of acquiring problem-solving skills from these so-called *video modeling examples* in which a (human) model explains and/or demonstrates how to perform a task on video (e.g., Groenendijk et al. 2013a, b; Hoogerheide et al. 2014; Van Gog et al. 2014). In addition to being effective for acquiring cognitive skills, observing video modeling examples has also been shown to enhance affective variables, such as students’ belief in their own ability to perform the modeled task at a certain level (i.e., self-efficacy; Bandura 1997; Schunk 1987).

When creating a video modeling example, an instructional designer is confronted with various design choices, which might affect learning, both cognitively as well as affectively. For instance, should the video present a natural task performance procedure, which might entail making and correcting errors (e.g., Groenendijk et al. 2013a; b), or a more didactical procedure that reflects how a student should ideally learn the skill (e.g., Hoogerheide et al. 2014; Simon and Werner 1996; Van Gog et al. 2014)? Another design consideration is whether the model should be (partly) visible in the video while explaining the task (e.g., Hoogerheide et al. 2014; Van Gog et al. 2014; Xeroulis et al. 2007), or whether only the model’s computer screen should be shown (e.g., McLaren et al. 2008; Van Gog 2011; Van Gog et al. 2009). If a form is chosen in which the model is visible, the question arises who the model should be in terms of expertise, age, background, and gender.

Because the widespread use of online video modeling examples is relatively recent, there is as of yet little empirical knowledge available to guide design choices. Recent studies have started to uncover effects of different ways of presenting the content in video modeling examples (e.g., to which degree the model should be visible; Hoogerheide et al. 2014; Van Gog et al. 2014). Potential effects of model characteristics that are unrelated to how the learning task is presented, such as gender, on the learning process and learning outcomes, have received little attention in recent research on video modeling examples. However, earlier research inspired by the model-observer similarity hypothesis (Schunk 1987, 1991), as well as recent research on pedagogical agents (e.g., Baylor and Kim 2004; Ozogul et al. 2013), suggests that similarity in factors such as gender may matter. Building on these findings, which will be reviewed below, the present study examined whether the effectiveness and efficiency of video modeling examples can vary as a function of the observer’s and model’s gender.

Model-observer similarity

The model-observer similarity hypothesis (Schunk 1987, 1991; see also the similarity-attraction hypothesis; Moreno and Flowerday 2006) states that model characteristics can matter when learning from modeling examples because the effectiveness of modeling is at least partly moderated by the degree to which observers perceive a model to be similar to them. Modeling evokes social comparison (Berger 1977; Johnson and Lammers 2012) and observing a model that successfully performs a task may lead observers to believe that they can perform the task as well, if they identify with the model (Bandura 1981; Schunk 1984). Moreover, an observer may be more attracted to and pay more attention to a model that is perceived as similar (Berscheid and Walster 1969).

As Schunk (1987) noted, “similarity serves as an important source of information for gauging behavioural appropriateness, formulating outcome expectations, and assessing one’s self-efficacy for learning or performing tasks” (p. 149). It is likely that particularly novice learners whose prior knowledge as well as self-efficacy and perceived competence are still low, are affected by model-observer similarity, as they are especially likely to engage in social comparison (Buunk et al. 2003). In other words, the higher the degree of similarity between observer and model, particularly when the observer is novice to the task at hand, the more cognitive outcomes of learning (e.g., performing the same or novel tasks) and affective aspects of the learning process (e.g., self-efficacy, perceived competence) may be enhanced.

With respect to those affective variables, self-efficacy is important because it influences factors such as academic motivation, study behaviour, and learning outcomes (Bandura 1997; Bong and Skaalvik 2003; Schunk 2001). Similarly, perceived competence, which is a related construct that reflects broader perceptions and knowledge (Bong and Skaalvik 2003; Hughes et al. 2011; Klassen and Usher 2010), also affects academic motivation and learning outcomes (Bong and Skaalvik 2003; Harter 1990; Ma and Kishor 1997). Moreover, when students’ confidence in their own capabilities increases, they tend to use more cognitive and metacognitive strategies irrespective of previous achievement or ability (Pajares 2006) and the willingness to invest mental effort in a task changes as well (Bandura 1977; Salomon 1983, 1984).

Gender can perhaps be expected to be the most important factor of model-observer similarity because gender is among the first things being noticed when interacting with others (Contreras et al. 2013). Schunk (1987), however, reported mixed results on both learning outcomes and self-efficacy in his review, and suggested that one possible explanation for these mixed findings might lie in the appropriateness of the modelled behaviour: students’ beliefs that a skill or behaviour is more appropriate for one of the genders may moderate effects of gender similarity. This might explain why Bandura et al. (1963) and Hicks (1965) found that for boys, observing a male model displaying aggressive behaviour towards a doll led to more imitative aggression than observing a female model. In contrast, no such effects were found for grade 4–6 students who observed a male or female model solving fraction problems (Schunk et al. 1987). Although mathematical tasks are typically more associated with males than females (Forgasz et al. 2004; Stewart-Williams 2002), young children do not yet seem to hold this association, which becomes stronger during adolescence (Steffens et al. 2010; see also Ceci et al. 2014). In other words, the 10 year olds in the study by Schunk et al. (1987) may have been too young to associate a mathematical task with gender.

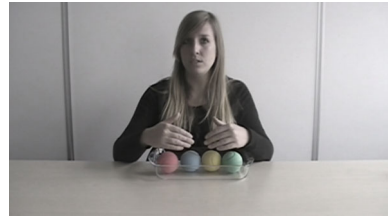
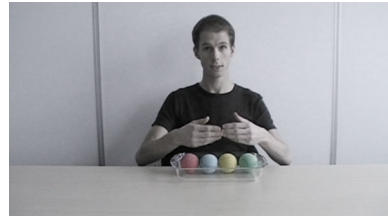
More recent studies also suggest mixed findings, however. Surprisingly in light of the above, a study with university students learning probability calculation with dynamic visualizations accompanied by a male or female model's narration showed that a female model was preferred and led to better learning outcomes than a male model (Linek et al. 2010). However, findings of Rodicio (2012) and Lee et al. (2007) suggest the opposite, namely that male narrations should be preferred. More specifically, Rodicio (2012) found that university students learned more about geology from dynamic visualizations with a male voice-over than a female voice-over, and Lee et al. (2007) found that for male students, a male computer-generated voice was more positively evaluated, trusted, and led to higher confidence levels than a female computer-generated voice. Note though, that in these studies, the model was not visible and therefore the cues available to make a social comparison may have been less strong compared to video modeling examples with a visible model (Hoogerheide et al. 2014).

Several animated pedagogical agent studies, in which a cartoon-like (humanoid) agent functions as a model or teacher, did show a preference for male agents, particularly for tasks that may be believed to be more appropriate for men. For instance, Moreno (2002) found that university students' knowledge about blood pressure was enhanced more after interacting with a male agent than a female agent. Arroyo et al. (2009) found that for secondary education and university students, a male agent led to more positive attitudes about mathematics and better learning outcomes. Furthermore, a study in educational technology found that male agents were evaluated as more interesting, intelligent, useful, and satisfactory than female agents (Baylor and Kim 2004). However, other research has shown that when learning an engineering task, often considered a stereotypically male domain in Western countries, interacting with a female model decreased women's beliefs about engineering stereotypes compared to interacting with a male agent (Rosenberg-Kima et al. 2008). Moreover, when given the choice, students tend to select an agent of the same gender (Ozogul et al. 2013).

In sum, the model-observer similarity hypothesis suggests that if one observes a same-gender model, affective and cognitive aspects of learning are more enhanced. More recent studies, particularly those with animated pedagogical agents, seem to suggest however, that for tasks that are more appropriate for males, male agents are preferred over female ones. Therefore, when it comes to video modeling examples, it is still an open question how gender affects learning.

The present study

The present study investigated whether it is more effective for male and female secondary education students to study video modeling examples depicting a same-gender model explaining and demonstrating a math task in terms of cognitive aspects of learning (i.e., learning and near transfer) and motivational aspects of learning (i.e., self-efficacy and perceived competence). In addition, the study measured cognitive load (i.e., effort investment) during the learning and test phase to investigate effects on the learning process and explored effects on judgment of learning accuracy and instruction evaluation. Female and male secondary education students learned how to solve probability calculation problems with replacement and order important by watching a video modeling example in which either a male (see Fig. 1) or a female (see Fig. 2) model explained and demonstrated the task. Both were instructed to wear a neutral, black t-shirt, and participated in an extensive practice training session to ensure that they showed the same behaviour

Fig. 1 Female model**Fig. 2** Male model

throughout the video (e.g., identical movements and gestures). An autocue was used to guarantee that the models gave the same explanation and spent the same amount of time on the steps shown in the video (and consequently on the video as a whole). After sufficient practice (as judged by the first author who was present at all times), the definitive recordings were created. Moreover, other factors that might affect perceived similarity were kept constant across conditions by selecting a young adult male and female Caucasian model (the majority of our participant population was Caucasian), who had a comparable educational background and were both in their early twenties. Therefore, we could be confident that effects (if any) would not be caused by differences in the content that was being presented.”

We firstly hypothesized that for male and female secondary education students who have little if any knowledge of solving probability calculation problems, it would be effective to study video modeling examples with both a male and female model, because research has consistently shown that example-based learning is very effective and efficient for novice learners (Atkinson et al. 2000; Renkl 2014; Sweller et al. 2011; Van Gog and Rummel 2010).¹ Thus, we expected high pretest to posttest performance gains (Hypothesis 1a) attained with a low to medium amount of effort investment during example study (Hypothesis 1b), while the amount of mental effort required to solve the test problems would decrease (Hypothesis 1c). Students’ self-efficacy and perceived competence were also expected to increase from pretest to posttest (Hypothesis 1d), since observing a model successfully explain and demonstrate a task has been shown to positively affect novices’ confidence in their own abilities (Bandura 1981; Hoogerheide et al. 2014; Schunk 1984).

The more interesting and open question was whether model-observer similarity would have an effect on cognitive and affective variables. In other words, would male and female students differ in the degree to which learning and transfer (Question 2a) and self-efficacy

¹ Note that for students who have some prior knowledge of solving probability calculation problems, examples would lose their effectiveness or may even start to hamper learning compared to practice problem solving (Kalyuga et al. 2001; this is an example of the expertise-reversal effect; see Kalyuga et al. 2003; Kalyuga and Renkl 2010).

and perceived competence (Question 2d) would be enhanced, mental effort invested in the test reduced (Question 2c), and in the degree that students invest mental effort during example study (Question 2b), depending on whether they observed a video modeling example that presented a male or a female model? Based on the model-observer similarity hypothesis, we could expect novice learners to identify more with a same-gender model relative to an opposite-gender one and therefore show cognitive and affective benefits when learning from a same-gender model (Schunk 1987). However, based on research with animated pedagogical agents (e.g., Arroyo et al. 2009; Moreno et al. 2002) and dynamic visualizations with a voice-over (Lee et al. 2007; Rodicio 2012), we might expect that novices benefit more from a male model than a female model because mathematical tasks are associated more with males than females (Forgasz et al. 2004; Stewart-Williams 2002). Moreover, because the confidence that learners have in their own capabilities is associated with how much effort they invest (Bandura 1977; Salomon 1983, 1984), differences in perceived capabilities across conditions could affect how much mental effort students invest during example study.

Because enhanced confidence can also be a negative outcome if it leads to overconfidence, which can be detrimental to students' regulation of their learning process (Dunlosky and Rawson 2012; Rhodes and Tauber 2011; Thiede et al. 2003), we instructed participants to predict their performance on the posttest. This judgment of learning was then matched to their actual performance to explore whether students' judgment of learning accuracy would depend on the gender of the model (Question 3). Because an increase in confidence leads to using more cognitive and metacognitive strategies (Pajares 2006), differences might especially arise if students differ in their self-efficacy and perceived competence depending on the model's gender.

Previous research has shown that instruction evaluation measures such as learning enjoyment may vary depending on the form of example-based instruction (Hoogerheide et al. 2014; see also Liew et al. 2013), and therefore we also explored effects on learning enjoyment and willingness to receive similar instruction in the future (Question 4) because these can be important indicators for the use of online examples during future self-study (Yi and Hwang 2003). Differences on these instruction evaluation measures might especially be dependent on whether there are differences in effort investment during example study because when practice effort decreases, enjoyment of practice may increase (Hyllegard and Bories 2009).

Method

Participants and design

The experiment had a 2×2 design, with Gender Model (Male vs. Female) and Gender Observer (Male vs. Female) as between-subject factors. Participants were 167 predominantly Caucasian secondary education students ($M^{\text{age}} = 13.50$, $SD = 0.59$; 80 male, 87 female) in their second year of general secondary education, which is the second highest level of secondary education in The Netherlands and has a total duration of 5 years. The students were randomly allocated to a female model (38 girls, 43 boys) or a male model (42 boys, 44 girls) condition. The experiment was conducted at a point in time at which probability calculation had not yet been taught in the curriculum.

Materials

All materials were presented using Qualtrics, which is a web-based survey software tool platform (<http://www.qualtrics.com>).

Video modeling example

Two video modeling examples were created, one with a female model (see Fig. 1) and one with a male model (see Fig. 2). Both models used the same example to address how one would ideally solve a probability calculation problem without replacement and with order important (i.e., an ideal procedure). The problem-state of this example was as follows: “The scouting staff brings 4 coloured balls for the cub scouts to play with. There is a red ball, a blue ball, a yellow ball, and a green ball. The cub scouts get to choose a ball one by one and prefer every colour equally. What is the chance that the red ball gets picked first and the green ball second?” The example then explained step-by-step how to solve this problem and briefly addressed what would happen in case it was an example of a probability calculation *with* replacement.

Both models were in their twenties, Caucasian, and wore a black neutral outfit while sitting behind a desk with the learning materials placed on the desk (i.e., the 4 different coloured balls and a platter; see Figs. 1 and 2). An autocue was used to guarantee that the models gave the same explanation and spent the same amount of time on the steps shown in the video (and consequently on the video as a whole). After sufficient practice, the definitive recordings were created. At the beginning of the video, all four items rested inside a platter. While explaining the models interacted with the learning materials to illustrate the problem-solving steps. For example, while explaining the first event—the chance that the red ball is picked first—both models picked up the red ball and held it in the air, after which they placed the red ball at the side of the platter.

Pretest and posttest

Two test versions were created that both consisted of six probability calculation problems. Within each test, four items measured learning (i.e., applying what has been learned to new tasks of the same type that have the same structural features but differ in surface features; solution procedures: $1/4 \times 1/3 = 1/12$, $1/11 \times 1/10 = 1/110$, $1/6 \times 1/5 \times 1/4 = 1/120$, and $1/8 \times 1/7 \times 1/6 \times 1/5 = 1/1680$) and two near transfer (i.e., applying what has been learned to new tasks of the same type that differ partly in structural features and differ in surface features; solution procedures: $1/6 \times 1/6 = 1/36$, $1/5 \times 1/5 = 1/25$). All problems required participants to fill in the correct answer and calculation. For example, one problem provided the following problem-state: “On a cold Sunday, a fisherman catches all the fish at from a small lake, one at a time. There are four fish swimming in the lake: a perch, a bream, a pike, and an eel. What is the chance that the bream is caught first, and the pike caught second?” The correct answer would be $1/4 \times 1/3 = 1/12$. The two test versions were parallel to each other, that is, the problems were structurally equivalent across both tests, but they differed in surface features (i.e., cover stories). The internal consistency (Cronbach’s alpha) of the pretest was .775 and of the posttest it was .741.

Mental effort

Effort investment was measured after every test item on the pretest and posttest and after watching the video modeling example using the subjective rating scale of Paas (1992),

which asks participants to indicate the effort they invested on a 9-point scale that ranges from (1) very, very low effort to (9) very, very high effort.

Self-efficacy and perceived competence

Self-efficacy was measured by asking participants to indicate on a 9-point scale, ranging from (1) very, very unconfident to (9) very, very confident, to what degree they believed that they mastered the skill probability calculation. This measure was adopted from Hoogerheide et al. (2014) and the phrasing of the questions is similar to Bandura (2006). To measure perceived competence, an adapted version of the scale by Williams and Deci (1996) was used. This scale consists of four items and asks participants to indicate to what degree the item applies to them, on a scale of 1 (not at all true) to 7 (very true). The item “I am able to achieve my goals in this course” was removed because this question did not apply to the present experiment, leaving the following three items: “I feel confident in my ability to learn this material”, “I am capable of learning the material in this course”, and “I feel able to meet the challenge of performing well in this course”. The word “course” was rephrased to “probability calculation problems”.

Judgment of learning

To measure judgment of learning, participants were asked on a scale of 0 to 6 to indicate how many probability calculation problems they expected to answer correctly if presented with a test.

Instruction evaluation

To investigate how participants experienced the video modeling example, they were asked after observing the video modeling example to indicate how enjoyable watching the video was and to what degree they would prefer to receive similar instruction in the future on a scale of 0 (lowest) to 10 (highest).

Procedure

The session took place at the computer lab of participants' school (ca. 45 min.). Before participants arrived, A4-papers were distributed over the computer lab containing the name of participants and a link to the Qualtrics questionnaire. This questionnaire presented 4 ‘question blocks’. Prior to each question block, participants received a plenary verbal instruction, after which they completed that specific question block. Question block 1 asked participants to fill in a general demographic questionnaire for which they received 90 s. Question block 2 contained the pretest (6 probability calculation problems and mental effort ratings), for which participants were instructed to not only write down their answer, but also the calculation. The remainder of question block 2 presented questions to measure self-efficacy and perceived competence. Question block 3 presented the video example (a YouTube video embedded in Qualtrics) followed by a mental effort rating and the instruction evaluation questions (i.e., learning enjoyment and willingness to receive similar instruction). Lastly, question block 4 first presented self-efficacy, perceived competence, and judgment of learning questions, followed by the posttest, which consisted of six probability calculation problems and mental effort ratings. Those that received version A

as the pretest now received version B as the posttest, and those that received B as the pretest now received version A.

Data analysis

A maximum of 8 points could be earned on both tests for the problems that measured learning, and a maximum of 4 for the problems that measured near transfer. Participants could earn 2 points per probability calculation problem: 1 point for a correct answer (0.5 for a partially correct answer; 0 for an incorrect or missing answer) and 1 point for the correct calculation (0 for an incorrect calculation). Both points were granted if participants wrote down the correct answer.

Averages were computed for participants' invested mental effort in completing the learning and near transfer test items, as well as the three items that measured perceived competence, on the pretest and posttest separately. We then computed a measure of judgment of learning accuracy by multiplying participants' judgment of learning (i.e., how many of the 6 problems participants predicted to correctly solve) by two and subsequently subtracting their actual test performance (range -12 to $+12$).

Four participants were removed from all analyses because of technical issues during the experiment (one participant) or too high prior knowledge as indicated by a total attained score greater than 50 % on the pretest (three participants). This left 163 participants in total, of which 87 observed a female model (43 female students, 44 male students) and 76 a male model (38 female students, 38 male students). One male student who observed a male model was removed from all test performance analyses and mental effort analysis (excluding invested mental effort in learning the video content) because he had to leave the experiment shortly after he started working on the posttest.

Results

The test performance and invested mental effort scores can be found in Table 1, the self-efficacy, perceived competence, and judgment of learning (accuracy) scores in Table 2, and the instruction evaluation ratings in Table 3.

Test performance

We tested Hypothesis 1a and Question 2a using a mixed ANOVA, with Test Moment (Pretest, Posttest) as within-subject factor and Gender Model (Female, Male) and Gender Observer (Female, Male) as between-subject factors. The scores obtained on the test items that measured learning showed a significant main effect of Test Moment, $F(1, 158) = 658.79$, $p < .001$, $\eta_p^2 = .807$. Participants performed significantly better on the Posttest ($M = 5.56$, $SD = 2.55$) than on the Pretest ($M = 0.30$, $SD = 0.64$). There was no main effect of Gender Model, $F(1, 158) = 1.71$, $p = .192$, nor of Gender Observer, $F < 1$. No other interaction effects were significant, $F_s < 1$. With regard to near transfer, the main effect of Test Moment was significant, $F(1, 158) = 154.96$, $p < .001$, $\eta_p^2 = .495$. Performance was significantly higher on the Posttest ($M = 1.85$, $SD = 1.66$) than on the Pretest ($M = 0.28$, $SD = 0.54$). The main effects of Gender Model and Gender Observer were not significant, $F_s < 1$. Furthermore, no other interaction effects were found ($F_s < 1$; Test Moment and Gender Observer, $F(1, 158) = 3.15$, $p = .116$).

Table 1 Mean (SD) of learning (range 0–12) and near transfer (range 0–4) scores and mental effort (range 1–9) per condition

	Male observer		Female observer	
	Male model (<i>n</i> = 42)	Female model (<i>n</i> = 44)	Male model (<i>n</i> = 43)	Female model (<i>n</i> = 38)
Test scores learning pretest	0.74 (1.52)	0.41 (0.68)	0.31 (0.57)	0.12 (0.45)
Test scores learning posttest	5.65 (2.69)	5.28 (2.59)	5.89 (2.35)	5.38 (2.67)
Test scores near transfer pretest	0.51 (1.00)	0.43 (0.47)	0.14 (0.25)	0.26 (0.73)
Test scores near transfer posttest	1.80 (1.70)	1.70 (1.63)	2.03 (1.77)	1.87 (1.59)
Mental effort learning pretest	4.82 (1.92)	5.08 (2.03)	4.95 (1.93)	5.05 (1.85)
Mental effort learning posttest	3.27 (1.51)	3.64 (1.78)	3.77 (1.26)	3.95 (1.45)
Mental effort near transfer pretest	5.28 (2.02)	5.24 (2.27)	4.97 (2.08)	4.92 (2.14)
Mental effort near transfer posttest	3.25 (1.79)	3.67 (1.90)	3.68 (1.46)	3.67 (1.68)
Mental effort during example study	2.12 (1.27)	2.97 (1.72)	3.11 (1.69)	2.77 (1.41)

Table 2 Mean (SD) of self-efficacy (range 1–9) and perceived competence (range 1–7) scores and judgment of learning (range 0–8) and judgment of learning accuracy (range –12 to 12) per condition

	Male observer		Female observer	
	Male model	Female model	Male model	Female model
Self-efficacy pretest	4.57 (2.06)	4.32 (2.03)	3.73 (1.88)	3.49 (1.75)
Self-efficacy posttest	6.12 (1.37)	5.87 (1.28)	5.39 (1.19)	5.28 (1.52)
Perceived competence pretest	3.67 (1.33)	3.63 (1.52)	3.06 (1.43)	3.29 (1.54)
Perceived competence posttest	5.30 (1.06)	4.72 (1.21)	4.79 (1.25)	4.63 (1.49)
Judgment of learning	4.12 (1.21)	4.05 (1.06)	3.70 (0.98)	3.86 (1.25)
Judgment of learning accuracy	0.79 (3.81)	1.13 (4.51)	–0.51 (3.56)	0.47 (3.75)

Table 3 Mean (SD) of learning enjoyment and willingness to receive similar instruction (ranges 0–10) scores per condition

	Male observer		Female observer	
	Male model	Female model	Male model	Female model
Learning enjoyment	5.36 (2.50)	4.47 (2.46)	4.36 (2.27)	4.98 (2.74)
Willingness to receive similar instruction	6.98 (2.63)	6.18 (2.74)	5.82 (2.45)	6.51 (2.72)

Mental effort

We tested Hypothesis 1b and Question 2b via a 2×2 ANOVA with Gender Model (Female, Male) and Gender Observer (Female, Male) as between-subject factors. There was no significant main effect of Gender Model on the invested mental effort *during example study*, $F < 1$, nor of Gender Observer, $F(1, 159) = 1.93$, $p = .167$. The interaction effect between Gender Model and Gender Observer was significant, $F(1, 159) = 5.03$, $p = .026$, $\eta_p^2 = .031$. To explore this interaction effect, we firstly compared the effects of Model Gender for each Observer Gender condition separately. There was only an effect of Model Gender for male students: it was less effortful for them to study an example by a male ($M = 2.24$, $SD = 1.28$) than a female model ($M = 2.97$, $SD = 1.72$), $t(74) = 2.12$, $p = .037$, $d = 0.486$ (medium effect; Cohen 1988). Secondly, we compared the effects of Observer Gender for each Model Gender separately. There was only an effect of Observer Gender for male model: observing a male model was more effortful for female students ($M = 3.11$, $SD = 1.69$) than male students ($M = 2.24$, $SD = 1.28$), $t(80) = 2.62$, $p = .011$, $d = 0.585$ (medium effect; Cohen 1988).

A mixed ANOVA with Test Moment (Pretest, Posttest) as within-subject factor and Gender Model (Female, Male) and Gender Observer (Female, Male) as between-subject factors was used to test Hypothesis 1c and Question 2c. The results showed a main effect of Test Moment on invested mental effort in completing the probability calculation problems that measured *learning*, $F(1, 158) = 75.90$, $p < .001$, $\eta_p^2 = .325$. Participants invested less effort to complete the problems that measured learning on the Posttest ($M = 3.71$, $SD = 1.49$) than on the Pretest ($M = 5.04$, $SD = 1.90$). There were no main effects of Gender Model and Gender Observer, $F_s < 1$. None of the interaction effects were significant ($F_s < 1$; Test Moment and Gender Observer, $F(1, 158) = 1.65$, $p = .201$).

For the average mental effort invested in completing the *near transfer* problems on the tests, a main effect of Test Moment was found, $F(1, 158) = 84.24$, $p < .001$, $\eta_p^2 = .348$. Again, participants invested less effort to complete the near transfer problems on the Posttest ($M = 3.60$, $SD = 1.69$) than on the Pretest ($M = 5.18$, $SD = 2.10$). There were no main effects of Gender Model and Gender Observer, nor were there significant interaction effects, $F_s < 1$.

Self-efficacy and perceived competence

Hypothesis 1d and Question 2d were tested using a mixed ANOVA with Test Moment (Pretest, Posttest) as within-subject factor and Gender Model (Female, Male) and Gender Observer (Female, Male) as between-subject factors. There was a main effect of Test Moment, $F(1, 159) = 113.26$, $p < .001$, $\eta_p^2 = .416$. Participants showed higher confidence in their abilities on the posttest ($M = 5.60$, $SD = 1.35$) than on the pretest ($M = 3.96$, $SD = 1.96$). There was no main effect of Gender Model, $F < 1$, but there was a main effect of Gender Observer, $F(1, 159) = 10.16$, $p = .002$, $\eta_p^2 = .060$, showing that male students ($M = 5.14$, $SE = 0.15$) were significantly more confident in their own abilities than female students ($M = 4.47$, $SE = 0.14$). None of the interaction effects were significant, $F_s < 1$. With regards to perceived competence, a main effect was found of Test Moment, $F(1, 159) = 191.72$, $p < .001$, $\eta_p^2 = .547$. Participants perceived their competence to be higher on the posttest ($M = 4.82$, $SD = 1.27$) than on the pretest ($M = 3.37$, $SD = 1.46$). There was no main effect of Gender Model, $F < 1$, nor of Gender Observer, $F(1, 159) = 3.14$, $p = .078$. There was no interaction between Gender Model and Gender

Observer, $F < 1$. The interaction between Test Moment and Gender Model was significant, $F(1, 159) = 4.81$, $p = .030$, $\eta_p^2 = .029$. A closer look at the data showed that observing a male model enhanced perceived competence more from pretest ($M = 3.32$, $SE = 0.16$) to posttest ($M = 4.98$, $SE = 0.14$), than observing a female model enhanced perceived competence improvement from pretest ($M = 3.46$, $SE = 0.16$) to posttest ($M = 4.67$, $SE = 0.14$). No other interaction was significant, $F_s < 1$.

Judgment of learning

We tested Question 3 via a 2×2 ANOVA with Gender Model (Female, Male) and Gender Observer (Female, Male) as between-subject factors. On the judgment of learning scores, there was no main effect of Gender Model, $F < 1$, nor of Gender Observer, $F(1, 159) = 1.90$, $p = .170$. The interaction between Gender Model and Gender Observer was not significant either, $F < 1$. With respect to the accuracy of the judgments of learning, no main effect of Gender Model was found, $F(1, 159) = 1.47$, $p = .227$, nor of Gender Observer, $F(1, 159) = 2.21$, $p = 1.39$. There was no significant interaction either, $F < 1$. One sample t -tests showed that for all four combinations of the 2×2 design, judgment of learning accuracy was not statistically different from zero, $t_s > .10$, indicating that male and female students were highly accurate in predicting their performance.

Instruction evaluation

The 2 (Gender Model: male, female) $\times 2$ (Gender Observer: male, female) ANOVA on how enjoyable watching the video examples was (Question 4) showed no main effects of Gender Model and Gender Observer, $F_s < 1$. There was, however, a significant interaction effect between Gender Model and Gender Observer, $F(1, 159) = 4.27$, $p = .040$, $\eta_p^2 = .026$. To explore this interaction effect, we firstly examined the effects of Model Gender for each Observer Gender condition separately, but none of the effects were significant. However, when the effects of Observer Gender were compared for each Model Gender separately, it was found that learning from a male model was significantly more enjoyable for male students ($M = 5.47$, $SD = 2.45$) than for female students ($M = 4.46$, $SD = 2.27$), $t(80) = 2.13$, $p = .036$, $d = 0.428$ (medium effect; Cohen 1988).

With respect to the degree to which participants preferred to receive instruction in a similar manner in the future, the same pattern of results was found as on the learning enjoyment question. Again, we found no main effect of Gender Model, $F < 1$, nor of Gender Observer, $F(1, 159) = 1.45$, $p = .230$, but there was a significant interaction between Gender Model and Gender Observer, $F(1, 159) = 4.02$, $p = .047$, $\eta_p^2 = .025$. When investigating the effects of Model Gender for each Observer Gender condition separately, no effects were found, but when we compared the effects of Observer Gender for each Model Gender separately, it was found that observing a male model caused male students ($M = 7.13$, $SD = 2.51$) to be significantly more positive about receiving similar instruction in the future than female students ($M = 5.82$, $SD = 2.45$), $t(80) = 2.39$, $p = .019$, $d = 0.528$ (medium effect; Cohen 1988).

Correlations were computed between the invested mental effort ratings for learning the video content and the two instruction evaluation questions because these constructs showed a very similar pattern of results (i.e., significant interaction effects showing a very similar

pattern). Surprisingly, effort invested in learning did not significantly correlate with how enjoyable watching the videos was, $r = -0.17$, $p = .831$, nor with the degree to which participants preferred to receive similar instruction in the future, $r = -0.10$, $p = .204$.

Discussion

This experiment investigated whether it would be more effective for secondary education students to study a video modeling example in which it was demonstrated how a math problem should be solved, with a same-gender model than an opposite gender model, as the model-observer similarity hypothesis would predict (Schunk 1987, 1991). With respect to cognitive aspects of learning, the results showed that, as expected, example study was effective for fostering learning and near transfer (i.e., high gains from pretest to posttest; Hypothesis 1a), regardless of the model's or the observer's gender. That is, gender did not affect the degree to which students improved their performance (Question 2a).

As one would expect, given the knowledge gains, the amount of mental effort students had to invest in solving the probability calculation problems decreased from pretest to posttest (Hypothesis 1c), and this effort reduction was not affected by gender either (Question 2c). In accordance with Hypothesis 1b, students invested a low/medium degree of effort during example study. There were, however, differences in the effort invested during example study as a function of model/observer gender (Question 2b). For male students it was less effortful to study a male model than a female model and observing a male model was less effortful for male students than female students (both medium effect sizes). This indicates that the learning process was more efficient for male students who observed a male model compared to female students and compared to male students observing a female model (see Van Gog and Paas 2008, for a discussion of efficiency in terms of the relation between mental effort and performance).

The affective variables of self-efficacy and perceived competence, both of which have been associated with better learning outcomes (Bandura 1997; Bong and Skaalvik 2003; Harter 1990; Ma and Kishor 1997; Schunk 2001), were also enhanced from pretest to posttest (Hypothesis 1d), although no effect of model-observer similarity was found (Question 2d). Male students did show higher self-efficacy than female students, which was, however, not associated with higher learning outcomes. This may be a consequence of the stereotypical perception that males are more competent in math than female students (Steffens et al. 2010), particularly among older students (Ceci et al. 2014), although typically very few, if any, actual performance differences are found between the genders (Hyde et al. 1990, 2008). Although the findings on perceived confidence combined with performance suggest that male students may have overestimated their performance, the judgment of learning accuracy results show that gender did not affect how accurate participants were at judging their own skills (Question 3). The stereotype that males are better than females at math could also explain why observing a male model enhanced perceived competence more from pretest to posttest than observing a female model, for both male and female students. Perhaps all students saw the male model as more of an expert than the female model (despite the fact that the content of the examples was fully identical) at this stereotypically male task. This is in line with findings on the effectiveness of animated pedagogical agents (e.g., Arroyo et al. 2009; Moreno et al. 2002).

We also found gender effects on both learning enjoyment and willingness to receive similar instruction in the future (Question 4), which may be indicators of how students might use such examples during future self-study online (Yi and Hwang 2003). Results

showed that studying a male model was more enjoyable for male students than female students and caused male students to be more positive about receiving similar instruction in the future than female students (both medium effects). While at first sight the pattern on the instruction evaluation questions and invested mental effort during learning appear identical, these measures did not correlate, indicating separate effects.

In sum, our results suggest that the gender of the model in video examples does not affect learning outcomes, but may influence affective aspects of learning. Notably, our study kept the content of the example videos entirely equal across conditions, so these effects only result from the differences in models. Effects on affective variables are important as these might influence students' self-study behaviour. So with video modeling examples being increasingly used in online learning environments, as they have become much easier to create and share, instructional designers creating such environments may want to consider the effects of model gender on male and female students' affect. Given that learning outcomes did not differ, but perceived competence was higher for students who studied a male video model, educational practitioners could give preference to designing and using video modeling examples with a male model when students learn a task that is associated more with males than females. However, given that students' gender interacted with the gender of the models on the evaluation of the instruction and on invested mental effort during example study, it is likely more advisable to create both a male and a female model version with identical content. These videos could be distributed to the learners via either an adaptive system that allocates students a male or female model depending on their own gender, or by allowing students to choose the model they want to learn from. The latter would have the added benefit of giving students an extra opportunity of regulating their own study behaviour, which should increase feelings of autonomy and thereby possibly raise their motivation and self-efficacy (Bandura 2001; Behrend and Thompson 2012; Clark and Mayer 2011; Ryan and Deci 2000). A similar argument has previously been made in the animated pedagogical agent literature (Ozogul et al. 2013). Because the gender of the model in a video modeling example does not seem to affect students' test performance, there seems to be no harm in providing students with the opportunity to choose the gender of their model, although future research should first examine whether our findings are replicated using tasks from other domains and over longer study periods.

Given that we used a single example, future research should also explore effects of the model's gender in relation to students' gender with multiple models to investigate whether the effects on affective variables would become stronger or weaker over time and if they would become stronger, whether they start to influence learning outcomes over time. It would also be interesting to compare effects of a set of examples by multiple male or female models to a mixed set of examples by male and female models.

Acknowledgments This research was funded by Kennisnet. The authors would like to thank Vincent van Dam, Arjan Bijleveld, and Bas Hellendoorn for facilitating this study. We also thank Eveline Stoker, Jan Engelen, and Chantal Hartgers for their help with conducting the study.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Arroyo, I., Woolf, B. P., Royer, J. M., & Tai, M. (2009). Affective gendered learning companion. In international conference on artificial intelligence and education. Brighton: IOS Press.
- Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, 70, 181–214. doi:[10.3102/00346543070002181](https://doi.org/10.3102/00346543070002181).
- Ayres, P., Marcus, N., Chan, C., & Qian, N. (2009). Learning hand manipulative tasks: When instructional animations are superior to equivalent static representations. *Computers in Human Behavior*, 25, 348–353. doi:[10.1016/j.chb.2008.12.013](https://doi.org/10.1016/j.chb.2008.12.013).
- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs: Prentice Hall.
- Bandura, A. (1981). Self-referent thought: A developmental analysis of self-efficacy. In J. H. Flavell & L. D. Ross (Eds.), *Cognitive social development: Frontiers and possible futures* (pp. 200–239). New York: Cambridge University Press.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs: Prentice Hall.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, 52, 1–26. doi:[10.1146/annurev.psych.52.1.1](https://doi.org/10.1146/annurev.psych.52.1.1).
- Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307–337). Greenwich, CT: Information Age Publishing.
- Bandura, A., Ross, D., & Ross, S. A. (1963). Vicarious reinforcement and imitative learning. *Journal of Abnormal and Social Psychology*, 67, 601–607.
- Baylor, A. L., & Kim, Y. (2004). Pedagogical agent design: The impact of agent realism, gender, ethnicity, and instructional role. In J. C. Lester, R. M. Vicari, & F. Paraguacu (Eds.), *Intelligent tutoring systems* (pp. 592–603). Berlin: Springer.
- Behrend, T. S., & Thompson, L. F. (2012). Using animated agents in learner-controlled training: The effects of design control. *International Journal of Training and Development*, 16, 263–283. doi:[10.1111/j.1468-2419.2012.00413.x](https://doi.org/10.1111/j.1468-2419.2012.00413.x).
- Berger, S. M. (1977). Social comparison, modeling, and perseverance. In J. M. Suls & R. L. Miller (Eds.), *Social comparison processes: Theoretical and empirical perspectives* (pp. 209–234). Washington, DC: Hemisphere.
- Bergmann, J., & Sams, A. (2012). *Flip your classroom: Reach every student in every class every day*. Eugene, OR: International Society for Technology in Education.
- Berscheid, E., & Walster, E. H. (1969). *Interpersonal attraction*. Reading, MA: Addison-Wesley.
- Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, 15, 1–40. doi:[10.1023/A:1021302408382](https://doi.org/10.1023/A:1021302408382).
- Buunk, B. P., Zurriaga, R., Gonzalez-Roma, V., & Subirats, M. (2003). Engaging in upward and downward comparisons as a determinant of relative deprivation at work: A longitudinal study. *Journal of Vocational Behavior*, 62, 370–388. doi:[10.1016/S0001-8791\(02\)00015-5](https://doi.org/10.1016/S0001-8791(02)00015-5).
- Ceci, S. J., Ginther, D. K., Kahn, S., & Williams, W. M. (2014). Women in academic science: A changing landscape. *Psychological Science in the Public Interest*, 15, 75–141. doi:[10.1177/1529100614541236](https://doi.org/10.1177/1529100614541236).
- Clark, R. C., & Mayer, R. E. (2011). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (3rd ed.). San Francisco: Pfeiffer.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction* (pp. 453–494). Hillsdale: Erlbaum.
- Contreras, J. M., Banaji, M. R., & Mitchell, J. P. (2013). Multivoxel patterns in fusiform face area differentiate faces by sex and race. *PLoS One*, 8, e69684. doi:[10.1371/journal.pone.0069684](https://doi.org/10.1371/journal.pone.0069684).
- Day, J. (2008). Investigating learning with web lectures (Doctoral dissertation). Available from Georgia Institute of Technology.
- Day, J., & Foley, J. (2006). *Evaluating web lectures: A case study from HCI*. Paper presented at the conference on human factors in computing systems, Montreal, Canada. Retrieved June 6, 2014 from <http://portal.acm.org/citation.cfm?id=1125493>.
- Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produced underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction*, 22, 271–280. doi:[10.1016/j.learninstruc.2011.08.003](https://doi.org/10.1016/j.learninstruc.2011.08.003).
- Forgasz, G. B., Leder, L. E., & Klosterman, P. (2004). New perspectives on the gender stereotyping of mathematics. *Mathematical Thinking and Learning*, 6, 389–420.

- Groenendijk, T., Janssen, T., Rijlaarsdam, G., & Van den Bergh, H. (2013a). Learning to be creative. The effects of observational learning on students' design products and processes. *Learning and Instruction*, 28, 35–47. doi:[10.1016/j.learninstruc.2013.05.001](https://doi.org/10.1016/j.learninstruc.2013.05.001).
- Groenendijk, T., Janssen, T., Rijlaarsdam, G., & van den Bergh, H. (2013b). The effect of observational learning on students' performance, processes, and motivation in two creative domains. *The British Journal of Educational Psychology*, 83, 3–28. doi:[10.1111/j.2044-8279.2011.02052.x](https://doi.org/10.1111/j.2044-8279.2011.02052.x).
- Harter, S. (1990). Causes, correlates, and the functional role of global self-worth: A life-span perspective. In R. J. Sternberg & J. Kolligian (Eds.), *Competence considered* (pp. 67–97). New Haven, CT: Yale University Press.
- Hicks, D. J. (1965). Imitation and retention of film-mediated aggressive peer and adult models. *Journal of Personality and Social Psychology*, 2, 97–100. doi:[10.1037/h0022075](https://doi.org/10.1037/h0022075).
- Hoogerheide, V., Loyens, S. M. M., & Van Gog, T. (2014). Comparing the effects of worked examples and modeling examples on learning. *Computers in Human Behavior*, 41, 80–91. doi:[10.1016/j.chb.2014.09.013](https://doi.org/10.1016/j.chb.2014.09.013).
- Hughes, A., Galbraith, D., & White, D. (2011). Perceived competence: A common core for self-efficacy and self-concept? *Journal of Personality Assessment*, 93, 278–289. doi:[10.1080/00223891.2011.559390](https://doi.org/10.1080/00223891.2011.559390).
- Hyde, J. S., Fennema, E., & Lamon, S. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, 107, 139–155. doi:[10.1037//0033-2909.107.2.139](https://doi.org/10.1037//0033-2909.107.2.139).
- Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A., & Williams, C. (2008). Gender similarities characterize math performance. *Science*, 321, 494–495. doi:[10.1126/science.1160364](https://doi.org/10.1126/science.1160364).
- Hyllegard, R., & Bories, T. L. (2009). Deliberate practice theory: Perceived relevance, effort, and inherent enjoyment of music practice: Study II. *Perceptual and Motor Skills*, 109, 431–440. doi:[10.2466/PMS.109.2.431-440](https://doi.org/10.2466/PMS.109.2.431-440).
- Johnson, C. S., & Lammers, J. (2012). The powerful disregard social comparison information. *Journal of Experimental Social Psychology*, 48, 329–334. doi:[10.1016/j.jesp.2011.10.010](https://doi.org/10.1016/j.jesp.2011.10.010).
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38, 23–32. doi:[10.1207/s15326985Sep3801_4](https://doi.org/10.1207/s15326985Sep3801_4).
- Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology*, 93, 579–588. doi:[10.1037//0022-0663.93.3.579](https://doi.org/10.1037//0022-0663.93.3.579).
- Kalyuga, S., & Renkl, A. (2010). Expertise reversal effect and its instructional implications: Introduction to the special issue. *Instructional Science*, 38, 209–215. doi:[10.1007/s11251-009-9102-0](https://doi.org/10.1007/s11251-009-9102-0).
- Klassen, R. M., & Usher, E. L. (2010). Self-efficacy in educational settings: Recent research and emerging directions. In T. C. Urdan & S. A. Karabenick (Eds.), *The decade ahead: Theoretical perspectives on motivation and achievement* (pp. 1–33). Bingley: Emerald Group Publishing Limited.
- Lee, K. M., Liao, K., & Ryu, S. (2007). Children's responses to computer-synthesized speech in educational media: Gender consistency and gender similarity effects. *Human Communication Research*, 33, 310–329. doi:[10.1111/j.1468-2958.2007.00301.x](https://doi.org/10.1111/j.1468-2958.2007.00301.x).
- Lenhart, A. (2012). *Teens and video: Shooting, sharing, streaming and chatting*. Retrieved December 11, 2012, from <http://www.pewinternet.org/Reports/2012/Teens-and-online-video/Findings.aspx>.
- Liew, T., Tan, S., & Jayothisa, C. (2013). The effects of peer-like and expert-like pedagogical agents on learners' agent perceptions, task-related attitudes, and learning achievement. *Educational Technology & Society*, 16, 275–286.
- Linek, S. B., Gerjets, P., & Scheiter, K. (2010). The speaker/gender effect: Does the speaker's gender matter when presenting auditory text in multimedia messages? *Instructional Science*, 38, 503–521. doi:[10.1007/s11251-009-9115-8](https://doi.org/10.1007/s11251-009-9115-8).
- Ma, X., & Kishor, N. (1997). Attitude toward self, social factors, and achievement in mathematics: A meta-analytic view. *Educational Psychology Review*, 9, 89–120. doi:[10.1023/A:1024785812050](https://doi.org/10.1023/A:1024785812050).
- McLaren, B. M., Lim, S., & Koedinger, K. R. (2008). When and how often should worked examples be given to students? New results and a summary of the current state of research. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the cognitive science society* (pp. 2176–2181). Austin: Cognitive Science Society.
- Moreno, R., & Flowerday, T. (2006). Student's choice of animated pedagogical agents in science learning: A test of the similarity-attraction hypothesis on gender and ethnicity. *Contemporary Educational Psychology*, 31, 186–207. doi:[10.1016/j.cedpsych.2005.05.002](https://doi.org/10.1016/j.cedpsych.2005.05.002).
- Moreno, K. N., Person, N. K., Adcock, A. B., Eck, R. N. V., Jackson, G. T., & Marineau, J. C. (2002). Etiquette and efficacy in animated pedagogical agents: The role of stereotypes. Paper presented at the AAAI Symposium on Personalized Agents, Cape Cod, MA.
- Ozogul, G., Johnson, A. M., Atkinson, R. K., & Reisslein, M. (2013). Investigating the impact of pedagogical agent gender matching and learner choice on learning outcomes and perceptions. *Computers & Education*, 67, 36–50. doi:[10.1016/j.compedu.2013.02.006](https://doi.org/10.1016/j.compedu.2013.02.006).

- Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive load approach. *Journal of Educational Psychology*, 84, 429–434. doi:[10.1037/0022-0663.84.4.429](https://doi.org/10.1037/0022-0663.84.4.429).
- Pajares, F. (2006). Self-efficacy during childhood and adolescence. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 339–367). Greenwich, CT: Information Age Publishing.
- Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, 38, 1–37. doi:[10.1111/cogs.12086](https://doi.org/10.1111/cogs.12086).
- Rhodes, M. G., & Tauber, S. K. (2011). The influence of delaying judgments of learning on metacognitive accuracy: A meta-analytic review. *Psychological Bulletin*, 137, 131–148. doi:[10.1037/a0021705](https://doi.org/10.1037/a0021705).
- Rodicio, H. G. (2012). Learning from multimedia presentations: The effects of graphical realism and voice gender. *Electronic Journal of Research in Educational Psychology*, 10, 885–906.
- Rosenberg-Kima, R. B., Baylor, A. L., Plant, E. A., & Doerr, C. E. (2008). Interface agents as social models for female students: The effects of agent visual presence and appearance on female students' attitudes and beliefs. *Computers in Human Behavior*, 24, 2741–2756. doi:[10.1016/j.chb.2008.03.017](https://doi.org/10.1016/j.chb.2008.03.017).
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 68–78. doi:[10.1037//0003-066x.55.1.68](https://doi.org/10.1037//0003-066x.55.1.68).
- Salomon, G. (1983). The differential investment of mental effort in learning from different sources. *Educational Psychologist*, 18, 42–50. doi:[10.1080/00461528309529260](https://doi.org/10.1080/00461528309529260).
- Salomon, G. (1984). Television is “easy” and print is “tough”: The differential investment of mental effort as a function of perceptions and attributions. *Journal of Educational Psychology*, 76, 647–658. doi:[10.1037//0022-0663.76.4.647](https://doi.org/10.1037//0022-0663.76.4.647).
- Schunk, D. H. (1984). Self-efficacy perspective on achievement behavior. *Educational Psychologist*, 19, 48–58. doi:[10.1080/00461528409529281](https://doi.org/10.1080/00461528409529281).
- Schunk, D. (1987). Peer models and children's behavioral change. *Review of Educational Research*, 57, 149–174.
- Schunk, D. H. (1991). *Learning theories: An educational perspective*. New York: Merrill.
- Schunk, D. H. (2001). Social cognitive theory and self-regulated learning. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 125–151). Mahwah, NJ: Erlbaum.
- Schunk, D. H., Hanson, A. R., & Cox, P. D. (1987). Peer-model attributes and children's achievement behaviors. *Journal of Educational & Psychology*, 79, 54–61. doi:[10.1037/0022-0663.79.1.54](https://doi.org/10.1037/0022-0663.79.1.54).
- Simon, S. J., & Werner, J. M. (1996). Computer training through behavior modeling, self-paced, and instructional approaches: A field experiment. *Journal of Applied Psychology*, 81, 648–659. doi:[10.1037//0021-9010.81.6.648](https://doi.org/10.1037//0021-9010.81.6.648).
- Spires, H. A., Hervey, L. G., Morris, G., & Stelpflug, C. (2012). Energizing project-based inquiry: middle grade students read, write, and create videos. *Journal of Adolescent & Adult Literacy*, 55, 483–493. doi:[10.1002/JAAL.00058](https://doi.org/10.1002/JAAL.00058).
- Steffens, M. C., Jelenec, P., & Noack, P. (2010). On the leaky math pipeline: Comparing implicit math-gender stereotypes and math withdrawal in female and male children and adolescents. *Journal of Educational Psychology*, 102, 947–963. doi:[10.1037/a0019920](https://doi.org/10.1037/a0019920).
- Stewart-Williams, S. (2002). Gender, the perception of aggression, and the overestimation of gender bias. *Sex Roles*, 46, 177–189. doi:[10.1023/A:1019665803317](https://doi.org/10.1023/A:1019665803317).
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.
- Thiede, K. W., Anderson, M. C. M., & Theriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, 95, 66–73. doi:[10.1037/0022-0663.95.1.66](https://doi.org/10.1037/0022-0663.95.1.66).
- Traphagan, T., Kucsera, J. V., & Kishi, K. (2010). Impact of class lecture webcasting on attendance and learning. *Educational Technology Research and Development*, 58, 19–37. doi:[10.1007/s11423-009-9128-7](https://doi.org/10.1007/s11423-009-9128-7).
- Van Gog, T. (2011). Effects of identical example-problem and problem-example pairs on learning. *Computers & Education*, 57, 1775–1779. doi:[10.1016/j.compedu.2011.03.019](https://doi.org/10.1016/j.compedu.2011.03.019).
- Van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior*, 25, 785–791. doi:[10.1016/j.chb.2009.02.007](https://doi.org/10.1016/j.chb.2009.02.007).
- Van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist*, 43, 16–26. doi:[10.1080/00461520701756248](https://doi.org/10.1080/00461520701756248).
- Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review*, 22, 155–174. doi:[10.1007/s10648-010-9134-7](https://doi.org/10.1007/s10648-010-9134-7).
- Van Gog, T., Verveer, I., & Verveer, L. (2014). Learning from video modeling examples: Effects of seeing the human model's face. *Computers & Education*, 72, 323–327. doi:[10.1016/j.compedu.2013.12.004](https://doi.org/10.1016/j.compedu.2013.12.004).

- Williams, G. C., & Deci, E. L. (1996). Internalization of biopsychological values by medical students: A test of self-determination theory. *Journal of Personality and Social Psychology*, 70, 767–779. doi:[10.1037/0022-3514.70.4.76](https://doi.org/10.1037/0022-3514.70.4.76).
- Xeroulis, G. J., Park, J., Moulton, C. A., Reznick, R. K., Leblanc, V., & Dubrowski, A. (2007). Teaching suturing and knot-tying skills to medical students: A randomized controlled study comparing computer-based video instruction and (concurrent and summary) expert feedback. *Surgery*, 141, 442–449. doi:[10.1016/j.surg.2006.09.012](https://doi.org/10.1016/j.surg.2006.09.012).
- Yi, M. Y., & Hwang, Y. (2003). Predicting the use of web-based information systems: Self-efficacy, enjoyment, learning goal orientation, and the technology acceptance model. *International Journal of Human-Computer Studies*, 59, 431–449. doi:[10.1016/S1071-5819\(03\)00114-9](https://doi.org/10.1016/S1071-5819(03)00114-9).